A new era of Innovation

Strengthening Security with AI

Tilting the Scales of Cyberspace in Favor of the Defender

Google

Table of Contents

03 Navigating the Opportunity of AI-Driven Defense
Foreword by **Heather Adkins**

04 The Defender's Dilemma
Data Visualization

06 The New Rules for AI Agents
Feature by **Royal Hansen**

10 Building Trust in AI
Case Study by **Evan Kotsovinos**

12 To Make AI Secure, Google Teams Built an AI That Attacks Itself
Feature

14 How to Build AI Security Resilience: 4 Attack Techniques
Case Study

16 Cybercrime: A Multifaceted National Security Threat
Feature

18 Responding to SharePoint Vulnerability CVE-2025-53770
Case Study

19 Beyond the Firewall
Perspectives on the invisible battles and shifting strategies of modern cybersecurity

Navigating the Opportunity of AI-Driven Defense

By Heather Adkins
VP, Security Engineering

For decades, the security community has operated under a heavy axiom known as the "Defender's Dilemma": to survive, we must be right 100% of the time, while an attacker only needs to get lucky once. It is a lopsided equation that has defined our industry. But with the arrival of AI, I believe we are standing on the precipice of a historical reversal. For the first time, we have the opportunity to tilt the scales of cyberspace decisively in favor of the defender.

It is true that AI has introduced a shift in the threat landscape. Adversaries are indeed leveraging these tools to move faster and automate complex tasks. But to view AI only as a weapon is to miss its far greater potential as a shield. The same speed and scale that empower attackers can be deployed by defenders to analyze vast telemetry, detect anomalies, and patch vulnerabilities at a velocity no human team could match. We are not just entering a new era of threats; we are entering the era of the automated immune system.

Consider our critical infrastructure—the digital backbone of our energy, finance, and healthcare systems. For years, securing these interconnected networks has been a challenge of scale and complexity. Today, AI offers us the ability to "shift left" fundamentally—identifying and fixing vulnerabilities in code before it ever ships, and monitoring operational networks with a level of vigilance that doesn't sleep. We are moving from a reactive posture, where we patch holes after a breach, to a proactive one, where we close the window of opportunity before an adversary can even pry it open.

I have spent my career in security engineering, from the early days of manual incident response to building planetary-scale defenses. If there is one lesson I hold to, it is that security is not a gatekeeper; it is an enabler. You cannot drive a car fast unless you trust the brakes; we cannot deploy AI to revolutionize healthcare or energy grids unless we trust the systems running them. In this sense, AI-driven security is the prerequisite for AI-driven innovation.

At Google, we see this not as a distant theory but as an immediate engineering reality. We are using AI to scale the instincts of our best defenders, turning manual insights into automated protections that benefit everyone. But we also know that "security by design" is a community effort. No single organization can solve this in a silo. The challenges we face—and the opportunities we have—require a collective baseline of defense where insights are shared and standards are raised together.

This magazine is an invitation to look past the binary noise of "doom" versus "utopia" and focus on the practical work ahead. AI is a tool, and like any powerful tool, its impact depends on the skill and intent of the hands that wield it. My hope is that you leave these pages recognizing that we are not helpless in the face of new threats. We are equipped. If we build responsibly and collaborate openly, we have the power to build a digital future that is not only faster and smarter, but fundamentally more secure than anything we have built before.

The technology to win this race is in our hands. Let's get to work.

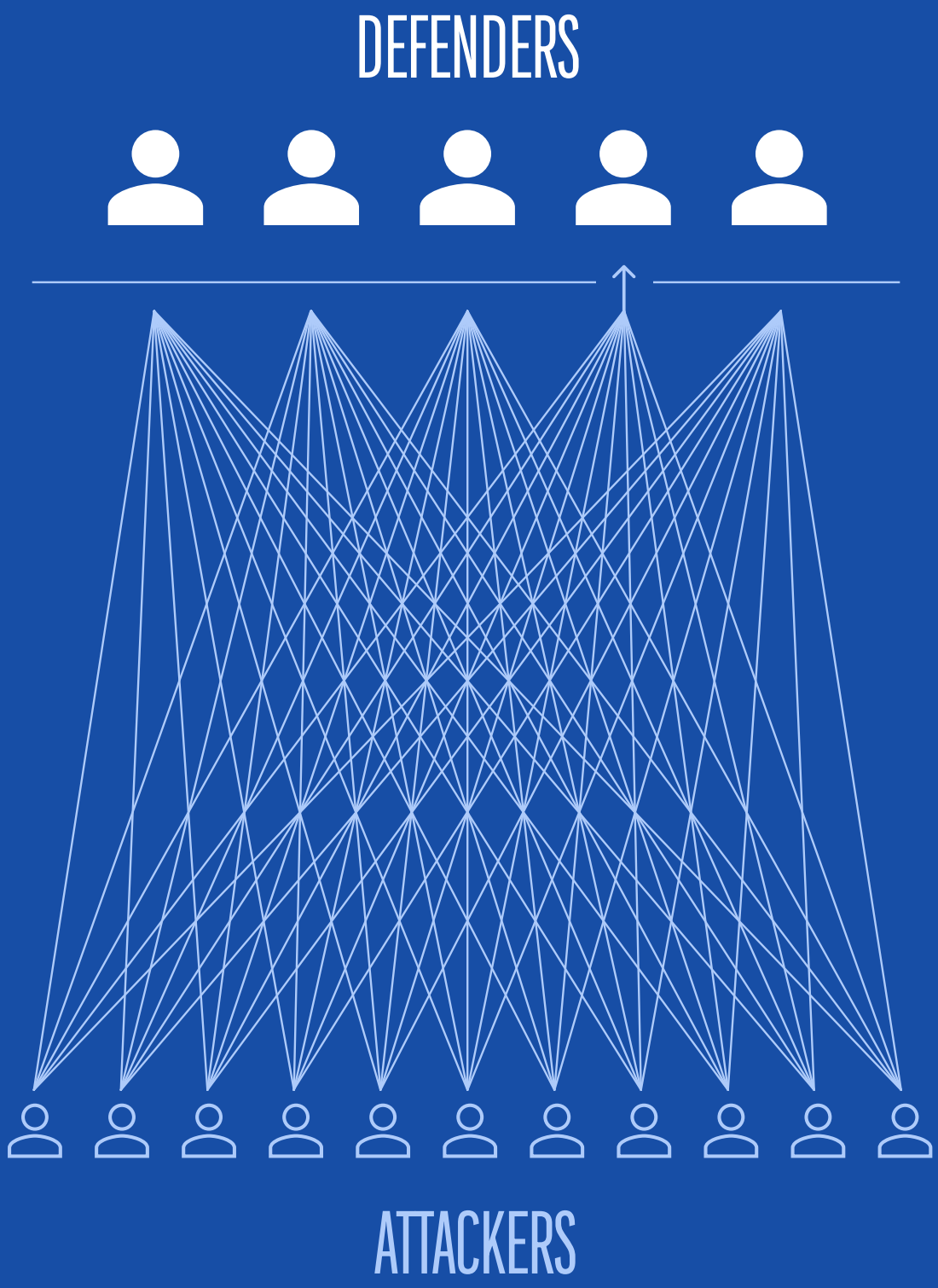
The Defender's Dilemma

The invention of the internet unlocked unprecedented innovation and economic opportunity.

But while the internet's core technologies fostered rapid innovation, interoperability, and the free flow of information, these technologies were not designed with security in mind.

The explosive growth of the digital domain on top of this foundation created an environment conducive to a wide range of malicious behaviors. Attackers possess inherent advantages in cyberspace: they can choose from a wide variety of targets and need only succeed once, while defenders must protect an increasingly complex terrain and need to be successful at all times. This dynamic, referred to as the "Defender's Dilemma," has plagued organizations and users for decades.

We believe AI affords the best opportunity to upend the Defender's Dilemma, and tilt the scales of cyberspace to give defenders a decisive advantage over attackers.


DEFENDERS
ATTACKERS

The New Rules for AI Agents

By Royal Hansen, VP, Engineering

As artificial intelligence moves from answering questions to taking action, Google's security framework reveals what it takes to keep autonomous systems safe—and accountable.

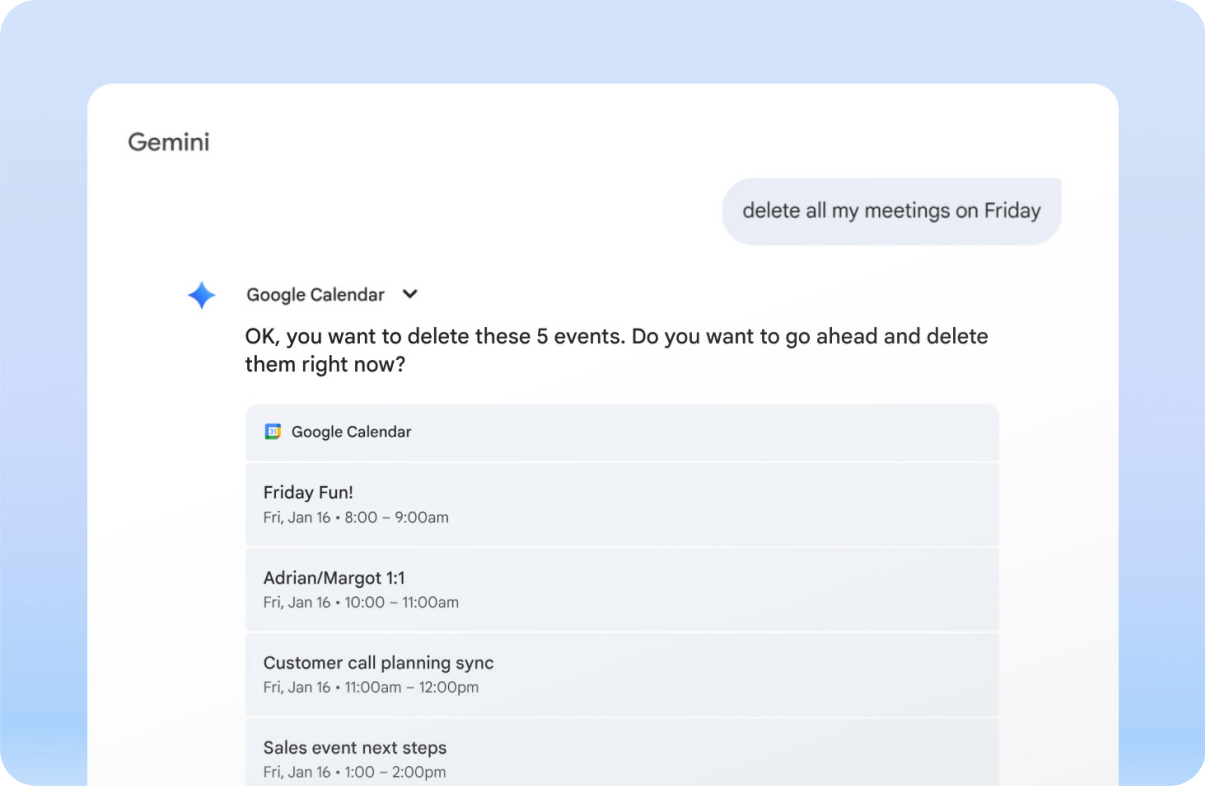
We have entered a new phase of AI. The era of one-way generative AI—systems that answer questions and generate text—is giving way to something fundamentally different. We think of agents as systems that combine the intelligence of advanced AI models with access to tools, so they can take actions on your behalf and under your control.

Agents can help with multi-step tasks like research, booking appointments, and even making purchases, always with your permission.

As agents become more helpful in everyday life, the industry is developing protocols, like the Universal Commerce Protocol (UCP), a new open standard for agentic commerce that works across the entire shopping journey - from discovery and buying to post-purchase support.

As the agentic ecosystem continues to evolve, it's essential to develop this technology safely and responsibly. At Google, three core principles—human oversight, innovation with control, and transparency—are critical for securing AI agents and the foundation for building trust in autonomous systems.

Human Oversight



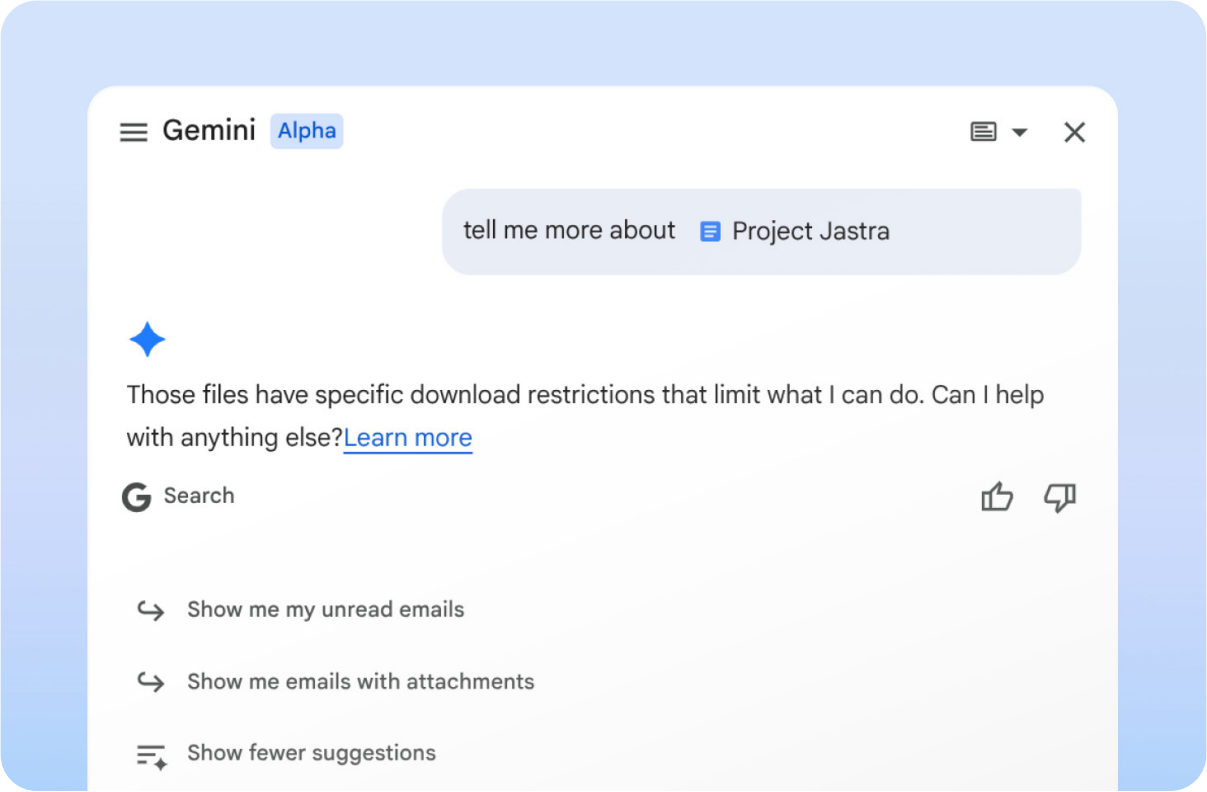
IN PRACTICE: *When the Gemini app prepares to delete entries in Google Calendar, it pauses before acting and will confirm it has your permission to delete.*

Agents operate in an open-ended world where not every action can be anticipated. Human confirmation must be given for critical or irreversible actions, such as authorizing significant financial transactions.

But every developer shouldn't have to build their own version of human oversight from scratch, creating custom systems to throttle output or decide when to pause for input. Google categorizes these decision points—when an agent should pause, when it should ask for confirmation, and when it should escalate to human judgment.

IN PRACTICE: *The Gemini "Human-In-The-Loop" framework enables Gemini in Workspace to request user feedback for certain actions to streamline the user experience. Users are now asked for confirmation on risky or final actions like deleting a calendar event.*

Innovation with Control



With agents, as capabilities increase, risks may increase as well. Setting clear limits on an agent's abilities is key to balancing usefulness with security. By carefully defining what an agent can and cannot do—like allowing an email assistant to manage messages but not financial accounts—can reduce the risk of serious harm if the agent makes a mistake. This ensures agents have only the capabilities and permissions for their intended purpose and cannot escalate their own permissions inappropriately.

But the level of access the agent should have needs to be decided. Should it inherit the user's full permissions, or operate with its own limited scope? The answer determines how authorization, authentication, and data access policies should apply. And the requirements vary dramatically by context.

For instance, an agent handling entertainment recommendations needs different access than one working in a hospital. These distinctions can't be captured in simple rules alone. Instead, they will require a paradigm for how agents will work alongside humans—one where each sector establishes its own specific security thresholds and development standards, as the risk profile varies significantly from one industry to the next.

IN PRACTICE: *When Gemini searches inside Gmail—looking for files that are confidential or need-to-know, for instance—it obeys the identity and access controls you'd expect, respecting permissions as it navigates.*

Transparency

Our final principle is transparency. Agent actions and planning must be observable. If we're going to trust these systems, we need to know what they're doing.

IN PRACTICE: *With Google Cloud Audit Logs and Gemini in Workspace, you can see which files an AI agent accessed, allowing you to quickly spot suspicious activity and build trust. This is a crucial element for incident handling and legal teams.*

The Path Forward

Google is embedding these principles into frameworks, protocols, and industry collaborations in partnership with industry leaders and researchers to scale responsible AI tools and practices. The security architecture we establish today will determine whether agents can safely take on the challenges that matter: energy, health, transportation.

However, the risks will be unique to each industry. The expertise of leaders in healthcare, in transportation, in defense will be critical to create the right foundation from the start, with the appropriate safeguards in place.

By working together, we can ensure AI is not only a force for incredible progress but is also developed and deployed in a way that is safe, trustworthy, and empowering for everyone.

Disclaimer: *This content is for illustrative purposes only and does not reflect the full scope of Google security protections. The Gemini app and user interface work differently for different scenarios, and we are continuously improving our defenses by adding new layers of protections.*

Building Trust in AI

By Evan Kotsovinos, VP, Privacy, Safety & Security

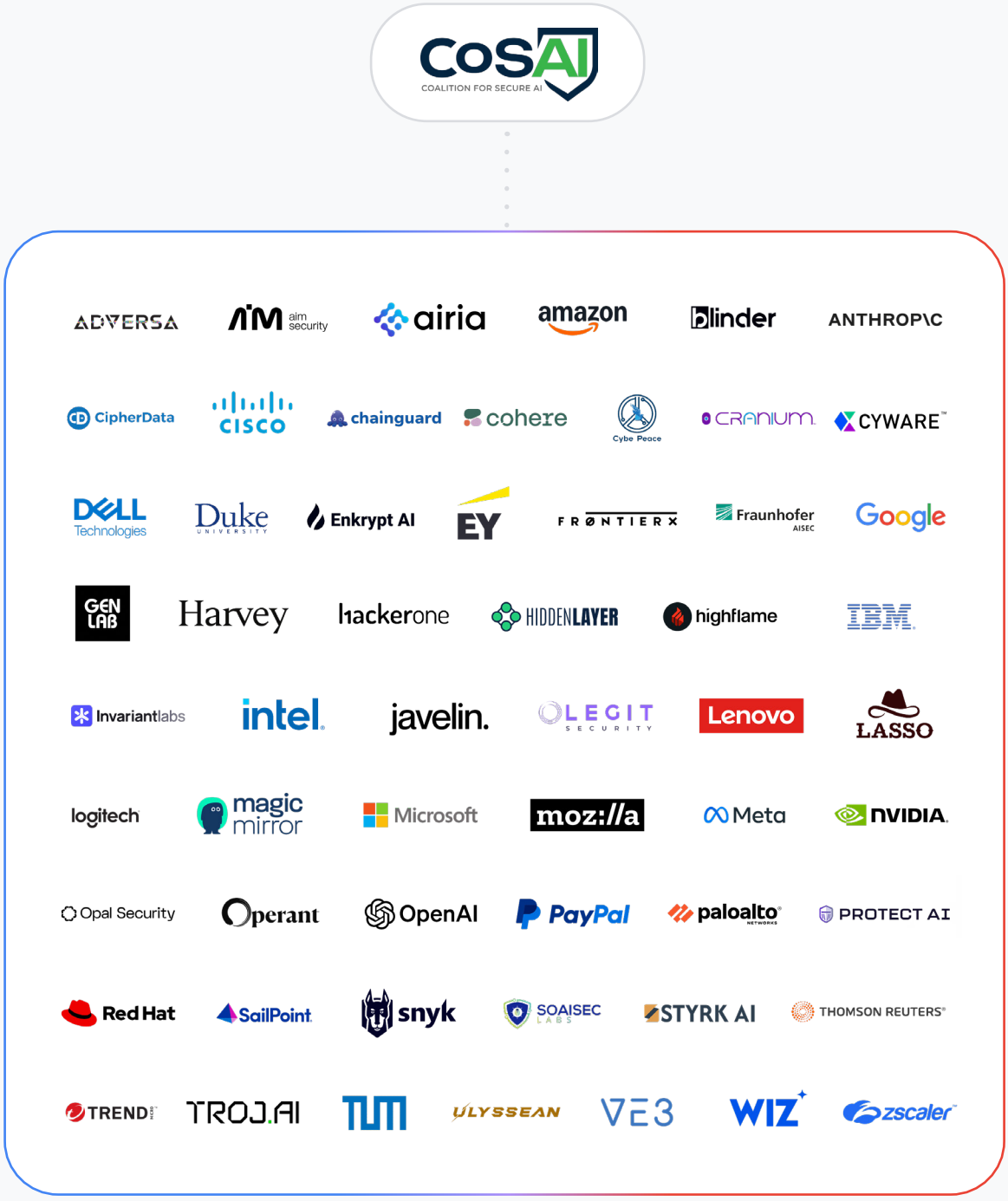
How Google and 'Coalition for Secure AI' are enhancing agent security

The evolution of AI presents incredible opportunities for our users, but also introduces new security challenges. In July 2024, Google, in collaboration with other industry partners, unveiled the 'Coalition for Secure AI (CoSAI)', an open ecosystem of leading AI and security experts working together on secure AI deployment, research, and product development.

Google's CoSAI participation remains at the forefront of emerging AI innovations, and specifically on a shared workstream focused on "Secure Design Patterns for Agentic Systems."

Leveraging decades of expertise in designing advanced security infrastructure, Google recently collaborated with CoSAI to develop these industry principles (human oversight, innovation with control and transparency) to build AI agents that are secure by design.

By collaborating to advance these security principles, CoSAI's 40+ leading AI and security firms are mitigating critical risks and paving the way for the safe and responsible evolution of agentic AI. This shared effort shows that collective action is key to realizing the transformative potential of AI agents while effectively managing their risks.


CoSAI
COALITION FOR SECURE AI
ADVERSA
aim security
airia
amazon
blinder
ANTHROP\C
CipherData
cisco
chainguard
cohere
Cybe Peace
CRANIUM
CYWARE
DELL Technologies
Duke UNIVERSITY
Enkrypt AI
EY
FRONTIERX
Fraunhofer AISEC
Google
GEN LAB
Harvey
hackerone
HIDDENLAYER
highflame
IBM
Invariantlabs
intel.
javelin.
LEGIT SECURITY
Lenovo
LASSO
logitech
magic mirror
Microsoft
moz://a
Meta
NVIDIA
Opal Security
Operant
OpenAI
PayPal
paloalto NETWORKS
PROTECT AI
Red Hat
SailPoint
snyk
SOAISEC LABS
STYRK AI
THOMSON REUTERS
TREND
TROJ.AI
TUM
ULYSSEAN
VE3
WIZ
zscaler

To Make AI Secure, Google Teams Built an AI That Attacks Itself

Imagine asking your AI agent to summarize your latest emails—a seemingly straightforward task. Large language models like Gemini are becoming increasingly capable of performing such tasks, pulling information from documents, calendars, and external websites. But what happens if one of those emails contains hidden, malicious instructions designed to trick the AI into sharing private data or misusing its permissions?

This scenario illustrates a growing cybersecurity challenge: indirect prompt injection. In this type of attack, adversarial commands are embedded within external data that the model retrieves and processes. When the AI can't distinguish between legitimate instructions and hidden manipulative commands, it risks acting on the attacker's intent instead of the user's.

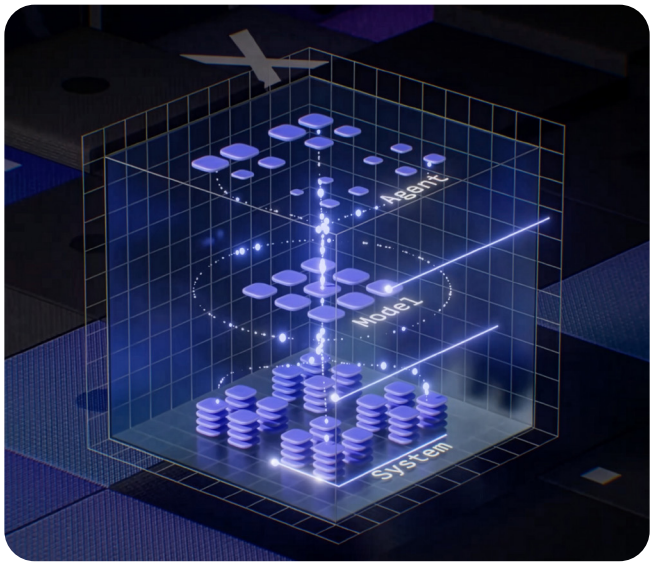
The urgency grows as AI systems take on more complex roles—retrieving external data, processing documents, and accessing sensitive information on behalf of users.

The Challenge

Current models don't perfectly distinguish between instructions from the user and content hidden in external data they retrieve. If successful, the model can deviate from the user's request and perform actions defined by the attacker instead.

Trying to find these vulnerabilities manually is slow and inefficient, especially as models evolve rapidly.

The Solution: Defense-in-Depth



Google is constantly working to make AI more secure for everyone, staying ahead of evolving threats like indirect prompt injection. Gemini helps block such attacks and alerts you with a warning.

Our "defense-in-depth" approach layers defenses to secure AI models for the long haul.

Google built an automated system to relentlessly probe Gemini's defenses. It works by having AI systems simulate attacks on other AI systems. Our framework continuously generates malicious prompts, some written in natural language, others using gibberish that somehow still manipulates the model and tests them against Gemini in realistic scenarios. The system learns from each attempt: when an attack fails, it tries variations; when one succeeds, it explores similar approaches. This attacking system creates new defense and attack techniques and those discoveries make Gemini more resilient.

Complementing this, Gemini's human red teams relentlessly probe Gemini's defenses to find and fix vulnerabilities before cybercriminals can exploit them.

The result is testing at a scale and speed that would be difficult for human testers alone, helping find vulnerabilities before they can be exploited in the real world.

Making Gemini More Resilient

This approach improved Gemini's ability to identify and ignore injected instructions—and crucially, without decreasing performance on normal tasks. Gemini 2.5 maintained competitive benchmark scores while becoming more resistant to attacks.

The results in practice:

- **If you ask Gemini to summarize emails** containing malicious prompts, those messages are excluded from the summary.
- **If you prompt Gemini to use file content,** it checks for suspicious patterns first and alerts you to potential security issues rather than following hidden instructions.
- **If retrieved data contains exfiltration attempts,** the model ignores them and continues executing only your legitimate request.

No single solution offers complete immunity. Even with improved defenses, determined attackers might find new vulnerabilities. The goal is to make attacks too expensive and complex for adversaries to execute.

Securing AI systems against evolving threats is an ongoing process. By layering defenses and learning constantly, we can enable models like Gemini to be both helpful and secure.

Disclaimer: *This content is for illustrative purposes only and does not reflect the full scope of Google security protections. The Gemini app and user interface work differently for different scenarios, and we are continuously improving our defenses by adding new layers of protections.*

How to Build AI Security Resilience: 4 Attack Techniques

To protect Gemini models from real-world threats, Google teams simulated common attack types.

The best way to strengthen an AI model's ability to fend off attackers is to simulate attacks and learn from the results. For example, with the goal of building Gemini 2.5's resilience against indirect prompt injections, Google DeepMind researchers implemented four different techniques: Actor-Critic, Beam Search, Tree of Attacks with Pruning (TAP), and Linear Generation attacks. They all simulate methods bad actors use to trick an AI model into following malicious orders.

Evaluating how Gemini responded to the attacks revealed weaknesses that Google's security researchers then addressed in Gemini 2.5's training data. The process was critical for preparing Gemini for a cyberthreat landscape that is shifting from manual human trickery to automated exploitation of system vulnerabilities. Google has deployed even stronger attack techniques to robustly test the latest Gemini 3 models.

Actor-Critic Attack

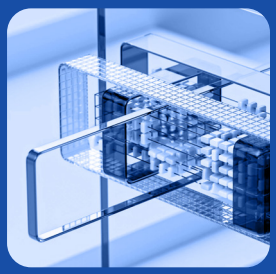
This technique involves reinforcement learning. An Actor model (the simulated attacker) generates adversarial prompts and tests them on the Target model (Gemini), which returns a probability score indicating how likely the attacker is to succeed. Based on this score, a Critic model suggests refinements to improve the attack, the Actor learns from the Critic's feedback, iteratively improving its attacks until it finds a successful trigger.

Beam Search Attack

This technique pursues multiple potential paths (beams) simultaneously to find ways around Gemini's security guardrails. Beam Search attacks often involve gibberish (i.e. strings of seemingly random characters, such as "!x#_&aa") that tricks the model into taking malicious action. These characters can be injected into HTML code or the metadata of a website where a human user won't notice them, but an AI model will still process them.

TAP Attack

TAP stands for "tree of attacks with pruning." The technique also involves Actor and Critic models, but they interact differently to create successful prompt injections. The Actor model generates multiple variations of a prompt (branches) to output to the Target model. The Critic prunes prompt candidates that are off-topic or ineffective, and gives suggested improvements to the Actor.



Linear Generation Attack

This technique uses AI to mimic persistent human attackers by rewriting successful prompt injections into diverse variants. Google DeepMind used this to create a massive dataset of triggers, which was integrated into Gemini's training to harden its defenses against real-world threats. Defending against complex indirect prompt injections requires constant vigilance. This effort is central to Google's agent security principles and our commitment to responsible AI.

Cybercrime: A Multifaceted National Security Threat

Cybercrime makes up a majority of the malicious activity online and occupies the majority of defenders' resources.

The Google Threat Intelligence Group (GTIG) is focused on identifying, analyzing, mitigating, and eliminating entire classes of cybersecurity threats against Alphabet, our users, and our customers.

In 2024, Mandiant Consulting responded to almost four times more intrusions conducted by financially motivated actors than state-backed intrusions. Despite this overwhelming volume, cybercrime receives much less attention from national security practitioners than the threat from state-backed groups. While the threat from state-backed hacking is rightly understood to be severe, it should not be evaluated in isolation from financially motivated intrusions.

Impact Over Motive

A useful way to think about the problem is through impact rather than motive. A hospital disrupted by a state-backed group using a wiper (malware) and a hospital disrupted by a financially motivated group using ransomware have the same impact on patient care. Likewise, sensitive data stolen from an organization and posted on a data leak site can be exploited by an adversary in the same way data exfiltrated in an espionage operation can be.

These parallels are especially troubling as cybercriminals expand their targeting of healthcare and critical infrastructure. Ransomware disruptions to energy systems (Colonial Pipeline in 2021, the Amsterdam–Rotterdam–Antwerp refining hub in 2022, Petro-Canada in 2023) have shown that financially motivated groups can pose threats that rival politically motivated attacks by undermining access to essential goods. A similar attack during a weather emergency or major crisis could have far more severe consequences.

The Economic Toll

In early 2024, a ransomware attack on a major U.S. healthcare payment processor disrupted billing systems across the country, with the parent company reporting $872 million in 'unfavorable cyberattack effects.' Two years earlier, Costa Rican President Rodrigo Chaves declared a national emergency caused by CONTI ransomware attacks against several government agencies. These intrusions caused widespread disruptions in government medical, tax, pension, and customs systems. With imports and exports halted, ports were overwhelmed, and the country experienced millions of dollars in losses. The remediation costs extended beyond Costa Rica; Spain supported the immediate response efforts, and in 2023, the U.S. announced $25 million in cybersecurity aid to Costa Rica.

Responding to a cybercrime incident can involve significant expenses: multi-million dollar ransom demands, loss of income due to system downtime, credit monitoring services for impacted clients, remediation costs, and fines. In the most extreme cases, these costs contribute to organizations ceasing operations or declaring bankruptcy.

The tactics are getting more aggressive and the extortion amounts are climbing higher. Extortion demands have reached as high as $50 million in some cases. Actual payments are typically lower, but the scale of these demands reflects how emboldened some criminal groups have become.

A Coordinated Response

Even if a single ransomware-as-a-service provider is taken down, many others are already in place to fill the gap. This resilient ecosystem means that while individual takedowns can disrupt particular operations and create temporary inconveniences for cybercriminals, these methods need to be paired with wide-ranging efforts to improve defense and crack down on these criminals' ability to operate.

A Comprehensive Approach is Required

While some welcome enhancements have been made in recent years, more must—and can—be done. We believe tackling this challenge requires a new and stronger approach recognizing the cybercriminal threat as a national security priority requiring international cooperation.

We urge policymakers to consider taking a number of steps:

- **Strengthen defenses against AI automated attacks.** Invest in detection capabilities that can identify AI-generated phishing content, social engineering tactics that bypass traditional filters, and "just-in-time" AI techniques, where malware queries language models mid-execution to evade detection.
- **Disrupt the cybercrime ecosystem's enabling infrastructure.** Target the maturing underground marketplace for illicit AI tools, and coordinate measures against malware developers, bulletproof hosting providers, and the financial intermediaries that enable these operations.
- **Prioritize international cooperation.** Cybercrime crosses borders; responses must as well. Expand information-sharing frameworks, support joint investigations, and contribute to multilateral initiatives aimed at dismantling transnational networks.
- **Build resilience in healthcare and critical infrastructure.** Healthcare's share of ransomware victims has doubled in three years, and energy systems remain frequent targets. Accelerate adoption of security best practices in these sectors, ensuring incident response plans account for attacks during crises when consequences would be most severe.

Responding to SharePoint Vulnerability CVE-2025-53770

When hackers saw a SharePoint weakness, they moved quickly.

During one weekend in July 2025, hundreds of organizations—including government agencies and private companies spread across multiple continents—saw their servers infiltrated in a global operation targeting Microsoft's SharePoint server software. About 10 days earlier, Microsoft had identified the security flaw and released patches to address it. Any organization running an on-premise (rather than cloud-hosted) SharePoint Server to store and manage documents was vulnerable.

The flaw, known as ToolShell and present in any SharePoint version released since 2016, was significant. It allowed hackers to gain unauthenticated access to everything on the server, including login credentials that could support ongoing stealth access. The potential victim list was massive: upwards of 9,000 servers owned and operated by banks, healthcare, and telecom companies and an array of government bodies, including the U.S. National Nuclear Security Administration.

Google Threat Intelligence Group (GTIG), which is focused on identifying and mitigating cybersecurity threats, was soon tracking the widespread SharePoint attackers and providing threat mitigation guidance to Google Threat Intelligence (GTI) subscribers.

Addressing the SharePoint vulnerability involved more than implementing Microsoft's updated, effective patches. If attackers gained access and stole server credentials, they could maintain access even if the ToolShell flaw was fixed. The solution GTIG recommended to SharePoint administrators? Change the SharePoint MachineKey so that any stolen keys would be rendered useless, fully evicting any attacker. GTIG also sent specific guidance for detecting and remediating ToolShell attacks to subscribers.

> Security isn't a destination to be arrived at, or a product of prescription – it is an active contest against motivated, adaptable adversaries. Because this threat is dynamic, it cannot be contained by static rules; defenses must be layered and resilient. Foundational approaches like building defense in depth technology, constantly monitoring for and remediating vulnerabilities, and disrupting the economics of exploitation are evergreen tools that can help skew the odds in favor of defenders.
>
> **Jasika Bawa**
> Group Product Manager, Chrome Security, Google

Beyond the Firewall

Perspectives on the invisible battles and shifting strategies of modern cybersecurity.

Security is often portrayed as a static shield or a lock on a digital door, but the reality of modern defense is far more dynamic. To cut through the noise, we invited ten leaders in cybersecurity to answer a single question:

What is something about security you wish more people knew?

Their reflections reveal a consensus that security is not a product you buy, but a process you build. Whether navigating the complexities of adversaries using AI or reinforcing the "paved roads" of software development, these leaders agree that the future of security belongs to the agile, the prepared, and the proactive.

> AI is moving fast, but a lot of the security basics, for instance core tenets like least privilege, defense in depth, zero trust and responsible disclosure, are as relevant as ever, and can be used to our advantage in today's complex security environment.
>
> **John "Four" Flynn**
> VP, Security and Privacy, Google DeepMind

AI WILL BE TRANSFORMATIVE FOR ADVERSARIES, BUT MANY OF THE ADVANTAGES WILL BE LESS EXTRAORDINARY THAN MOST IMAGINE.

One of the greatest advantages AI will give criminals and state actors will be simply scaling up their operations. An adversary using this technology will be able to hit more targets effectively than previously possible. This scale will also translate to speed and persistence. Expect them to move faster on n-days or to test your network until an opportunity presents itself. Notably, AI may be the best option for defenders to respond quickly and at scale to threats like this.

Sandra Joyce
VP, Google Threat Intelligence

“WE ARE ON THE PRECIPICE OF A SEISMIC SHIFT IN CYBER OFFENSE. THREAT ACTORS OF ALL TYPES WILL LEVERAGE AI ACROSS THE FULL SPECTRUM OF THEIR OPERATIONS TO AUTOMATE AND SCALE EXPLOITATION AND DISRUPTION. A CATEGORY 5 HURRICANE IS COMING—AND DEFENDERS WHO DON’T ADOPT AI JUST AS AGGRESSIVELY WILL BE FLATTENED.”

Dmitri Alperovitch
Chairman, Silverado Policy Accelerator

Anyone building Agentic AI should treat security as a customer–supplier partnership. Because agents act as software identities on a user's behalf, they require the same fundamentals as secure human collaboration: strong authentication, clear authorization, and least-privilege access tied to a specific goal. Both sides must jointly define identity, permissions, policy enforcement, and guardrails for personal and confidential data. Controls should be reviewed continuously, not assessed once. Third-party risk must also expand to cover data handling, retention, model-training use, continuous monitoring, audit logs, and coordinated incident-response processes that both parties trust.

Scott Moser
SVP, Secure Cloud Infrastructure and CISO, Sabre

Even the most sophisticated attackers are ruthlessly efficient: they choose the cheapest and fastest way in. And, we often make it too easy. The Change Healthcare attack paralyzed the U.S. medical system not with a zero-day exploit, but via a single server missing MFA. China's telecom breaches exploited similar gaps. This is why I'm hopeful that AI-generated digital twins will be a game changer. By simulating common attacks and quantifying the financial cost of disruption, they can turn abstract cyber risk into a bottom line, incentivizing leaders to fund the fixes needed.

Anne Neuberger
Distinguished Fellow, Stanford University, Senior Advisor, a16z, former Deputy National Security Advisor

WE ALWAYS HEAR HOW SECURITY MUST 'SHIFT LEFT.' SECURITY ISN'T JUST A FINAL LAYER WE WRAP AROUND A PRODUCT; IT'S THE FOUNDATION THAT MAKES AI POSSIBLE.

I wish more people understood that security is actually the accelerator for innovation – not the barrier. By building 'paved roads' – the essential guardrails – into the core of our AI models from day one, we create the trust necessary for businesses to move faster. At Workday, we believe that 'security by design' ensures users never have to choose between cutting-edge AI and data integrity.

Bill Shields
CISO, Workday

While pessimism around AI and cybersecurity is common, I think 2026 will be a positive turning point. Security programs are largely evolving beyond rigid compliance frameworks toward proactive attack simulation as a strategy. AI tooling is a huge catalyst, allowing us to identify and remediate material issues rather than obsessing over minor process flaws. We are also seeing AI development tools solve complex software security problems at scale. Let's move past the fear and embrace automation to solve software security and focus our programs on attack simulation to improve!

John Rogers
CISO, MSCI

Cybersecurity is ultimately a leadership issue. It isn't a product you buy; it's a culture you build. Technology matters, but outcomes depend on how well people, processes, and priorities are aligned. The strongest security environments are built deliberately, over time, with clear ownership and accountability. Most cyber failures don't start with sophisticated attacks—they start with the basics being ignored. Strong security is built through people, process, and accountability, applied consistently over time. When leaders treat it as an operational priority—not a compliance exercise—risk drops and resilience actually improves.

Andy Boyd
Operating Partner at AE Industrial Partners and CEO of REDLattice/Paragon

Security is often mistakenly viewed as a static defense, such as a simple door lock. However, defending users from modern threats demands a dynamic and constantly evolving infrastructure. Relying on outdated or legacy systems to combat next-generation threats and automated abuse is a recipe for failure. This challenge is magnified by the disruptive force of AI, which attackers are leveraging in both predictable and unexpected ways. Effective security today necessitates modernizing our core architecture to be just as sophisticated and agile as the attacks we are working to neutralize.

Eduardo Tejada
VP, Core User Protection, Google


Google is committed to helping strengthen security and resilience globally.
Industry needs to do their part, with products that are secure-by-default and enabled by transparency and investments to support a safe, open internet that benefits everyone. Meeting this moment also requires organizations to move quickly and at scale, which will require strong partnerships between all stakeholders in the security ecosystem.
Learn more, at publicpolicy.google/security

Notes

Notes